CST 361/408 Deep Learning Algorithms	Spring 2022
Tu, Th 2:00pm-3:15pm, AUD 516	**HW 6**

Each exercise is 2pts

1. Ch. 7 sec. 10 ex. 4: Consider a recurrent network in which the hidden states have a dimensionality of 2. Every entry of the 2 ×2 matrix W_{hh} of transformations between hidden states is 3.5. Furthermore, sigmoid activation is used between hidden states of different temporal layers. Would such a network be more prone to the vanishing or the exploding gradient problem?
   Hint:
   a. Look for guidance at slides 16,17,18 lecture 9-1.
   b. Note that whenever on forward propagation we multiply by matrix W on backpropagation we multiply by matrix W^T
2. Instead of gradient of sigmoid use maximum value of this gradient Suppose we use echo state network with 5 hidden time layers that have sigmoid activation. Suppose we know that there is no vanishing gradient problem in our network due to nature of inputs, but there may be exploding gradient problem. What should be maximal spectral radius of the weight matrices that we can set to avoid getting gradient of final cell being blown up by more than 8 times?
   Hint: use bound derived in 1.
3. As can be seen from slide 3 lecture 8.2. RNN is composed of RNN cells similarly to LSTM or GRU. To see that consider formulas on slide 3 lecture 8.2 as defining one cell. Show the computational graph of forward computation of one RNN cell.